

REMARKS

This Preliminary Amendment is filed in order to facilitate processing of the above-identified application and is filed in response to the Office Action dated June 17, 2003, in which the Examiner rejected claims 1-5, 13, 16, 19 and 20 under 35 U.S.C. § 102(b) and rejected claims 6-12, 14, 15, 17 and 18 under 35 U.S.C. § 103.

Applicants would like to thank the Examiner for the telephone interview on September 15, 2003.

As indicated above, claims 1 and 16 have been amended to make explicit what is implicit in the claims. It is respectfully submitted that the amendment is unrelated to a statutory requirement for patentability.

Claim 1 claims a method for extracting information from a natural language text corpus based on a natural language query and claim 16 claims a system for extracting information. The method and system include analyzing the natural language text corpus with respect to surface structure of word tokens and surface syntactic roles of constituents. The analyzed natural language text corpus is indexed and stored. A natural language query is analyzed with respect to surface structure of word tokens and surface syntactic roles of constituents. A number of surface variants of the analyzed natural language query are created by replacing word tokens of the natural language query, and for at least one surface variant by rearranging word tokens of the natural language query, in such a way that the number of surface variants are equivalent to the natural language query with respect to lexical meaning of word tokens and surface syntactic roles of constituents. The number of

surface variants and the analyzed natural language query are compared with the indexed and stored analyzed natural language text corpus. From the indexed and stored analyzed natural language text corpus an extraction is made on each portion of text comprising a string of word tokens that matches any one of the surface variants or the analyzed natural language query.

Through the method and structure of the claimed invention a) creating a number of surface variants of the analyzed natural language query by replacing word tokens of the natural language query and for at least one surface variant by rearranging word tokens of the natural language query in such a way that the number of surface variants are equivalent to the natural language query with respect to lexical meaning of word tokens and surface syntactic roles of constituents, and b) comparing both the number of surface variants and the analyzed natural language query with the indexed and stored natural language text corpus, as claimed in claims 1 and 16, the claimed invention provides a method and system of extracting information in which the number of matches is increased relative to what it would be if matches were only verbatim searched. Since a number of surface variants and the original query are used in the matching process, catching linguistic variations present in the text corpus to be searched can be obtained without the need for a complex matching criteria such as when a regular expression is used. This allows for a more straightforward matching process where each surface variant is compared to the text corpus. The prior art does not show, teach or suggest the invention as claimed in claims 1 and 16.

Claims 1-5, 13, 16, 19 and 20 were rejected under 35 U.S.C. § 102(b) as being anticipated by *Julliard* (European Patent Application No. 0 886 226).

Julliard appears to disclose a method of searching for information in a text database, comprising: (a) receiving at least one user input, the user input(s) defining a natural language expression including one or more words, (b) converting the natural language expression to a tagged form of the expression, the tagged form including said one or more words and, for the or each word, a part-of-speech tag associated therewith, (c) applying to the tagged form one or more grammar rules of the language of the natural language expression, to derive a regular expression, and (d) analyzing the text database to determine whether there is a match between said regular expression and a portion of said text database. (col. 1, lines 31-44) The linguistic search techniques provide a new way to search for information in a text database. They enable users to find portions of a text which match multiword expressions given by the user. Matches include possible variations that are relevant with the initial criteria from a linguistic point of view including simple inflections like plural/singular, masculine/feminine or conjugated verbs and even more complex variations like the insertion of additional adjectives, adverbs, etc. in between the words specified by the user. This technique can complement conventional full text search engines by reducing the number of retrieved documents that are inconsistent with the query. (col. 2, lines 32-44) Figure 2 is a schematic flow diagram of the steps performed in carrying out a linguistic search. Initially (step s1), the user specifies the multiword expression he is looking for. Next, at step s2, the expression is then sent to the tagger (or

disambiguator). The tagger (or disambiguator) does two things--(1) reduce each word to its root form, and (2) determine the part-of-speech of each word. (col. 3) Once the tagged form 50 has been obtained, it is then simplified, at step s3: because it is desired that the linguistic search process retrieves all possible inflections of a word each tag is first reduced to its syntactic category. The process continues at step s4, in which the simplified tagged form 51 is operated on. Given the grammar of a language it is possible to determine what kind of variations a multiword expression can undergo without changing its initial meaning. (col. 4) The grammar rules expressed in step s4 are coded in a regular expression and matched against the simplified tagged form 51 of the user query. If one of those rules matches, then the simplified tagged form 51 of the user query transforms into a complex regular expression representing the grammar variations. The matching regular expression 52 is then processed further at step s5. Once the final regular expression 52 has been generated it is matched against the tagged version of the corpus. (col. 5) Step s6 is performed after the regular expression has been matched against the tagged version of the corpus. As mentioned above, the Perl (or awk) regular expressions mechanism can tell the user what string matches but also where this string is located in the text. However, because the regular expression matching is done on the tagged version of the corpus, the positioning information is not suitable for the original text. As a consequence, if it is desired to highlight the matches a way must be provided to go from the offset in the tagged text into the actual offset in the original text. Currently, this is made via a simple offset table built during the corpus tagging. (col. 6)

Thus, *Julliard* merely discloses coding grammar rules into a regular expression and as matched against a simplified tagged form. When a match occurs, the simplified tagged form of the user query is transformed into a complex regular expression representing the grammar variations. The final regular expression once generated is then matched against the tagged version of the corpus (col. 4, line 56 through col. 5, line 10). Thus, *Julliard* merely discloses creating one regular expression having a complex form representing the grammar variations. Nothing in *Julliard* shows, teaches or suggests creating a number of surface variants by replacing word tokens of the natural language query and for at least one surface variant by rearranging word tokens of the natural language query in such a way that the number of surface variants are equivalent to the natural language query with respect to lexical meaning of word tokens and surface syntactic roles of constituents as claimed in claims 1 and 16. Rather, *Julliard* merely discloses generating a regular expression representing grammar variations which is then matched against the tagged version of the corpus.

Additionally, *Julliard* merely discloses comparing the final regular expression against the tagged version of the corpus (column 5, lines 7-10). Nothing in *Julliard* shows, teaches or suggests comparing a) the number of surface variants and b) the analyzed natural language text with the indexed and stored analyzed natural language text corpus as claimed in claims 1 and 16. Rather, *Julliard* merely discloses matching a final regular expression against the tagged version of the corpus.

Since nothing in *Julliard* shows, teaches or suggests a) creating a number of surface variants by replacing word tokens of the natural language query and for at least one surface variant rearranging word tokens and b) comparing both the number of surface variants and the analyzed natural language query with the indexed and stored natural language text corpus as claimed in claims 1 and 16, it is respectfully requested that the Examiner withdraws the rejection to claims 1 and 16 under 35 U.S.C. § 102(b).

Claims 2-5, 13, 19 and 20 depend from claims 1 and 16 and recite additional features. It is respectfully submitted that claims 2-5, 13, 19 and 20 would not have been anticipated by *Julliard* within the meaning of 35 U.S.C. § 102(b) at least for the reasons as set forth above. Thus, it is respectfully requested that the Examiner withdraw the rejection to claims 2-5, 13, 19 and 20 under 35 U.S.C. § 102(b).

Claims 6-12 and 18 were rejected under 35 U.S.C. § 103 as being anticipated over *Julliard* in view of *Arampatzis et al.* ("An Evaluation of Linguistically-motivated Indexing Schemes," hereinafter referred to as *Arampatzis-Indexing*). Claims 14, 15 and 17 were rejected under 35 U.S.C. § 103 as being unpatentable over *Julliard* in view of *Arampatzis et al.* ("Linguistically-motivated Information Retrieval," hereinafter referred to as *Arampatzis-Retrieval*).

Applicants respectfully traverse the Examiner's rejection of the claims under 35 U.S.C. § 103. The claims have been reviewed in light of the Office Action, and for reasons which will be set forth below, it is respectfully requested that the Examiner withdraw the rejection to the claims, and allows the claims to issue.

As indicated above, since nothing in *Julliard* shows, teaches or suggest the primary feature as claimed in claims 1 and 16, it is respectfully requested that the combination with the secondary references to *Arampatzis-Indexing* and *Arampatzis-Retrieval* would not overcome the deficiencies of the primary reference. Furthermore, neither *Arampatzis-Indexing* or *Arampatzis-Retrieval* shows, teaches or suggests that any variants of a query should be created, which variants are equivalent to a query with respect to the lexical meanings of word tokens and surface syntactic roles of constituents. Therefore, it is respectfully submitted that the combination of the primary reference and the secondary references will not overcome the deficiencies of the primary reference. It is respectfully requested that the Examiner withdraw the rejection to claims 6-12, 14, 15, 17 and 18 under 35 U.S.C. § 103.

As indicated above, new claims 21-28 have been added. It is respectfully submitted that these claims are also in condition for allowance.

Thus, it now appears that the application is in condition for reconsideration and allowance. Reconsideration and allowance at an earlier date are respectfully requested.

If for any reason the Examiner feels that the application is not now in condition for allowance, it is respectfully requested that the Examiner contact, by telephone, the applicants' undersigned attorney at the indicated telephone number to arrange for an interview to expedite the disposition of this case.

In the event that this paper is not timely filed within the currently set shortened statutory period, applicants respectfully petition for an appropriate extension of time. The fees for such extension of time may be charged to our Deposit Account No. 02-4800.

In the event that any additional fees are due with this paper, please charge our Deposit Account No. 02-4800.

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

Date: October 17, 2003

By: 

Ellen Marcie Emas

Registration No. 32,131

P.O. Box 1404
Alexandria, Virginia 22313-1404
(703) 836-6620